



# Capítulo 12

---

*Análisis de Supervivencia*



# Capítulo 12

## Análisis de Supervivencia

---

### 1. Introducción

---

El **análisis de supervivencia** (*survival analysis*) persigue estimar la probabilidad de que ocurra un determinado suceso en función del tiempo; o lo que es lo mismo, describir la evolución de la supervivencia a lo largo del tiempo. Este análisis se aplica, pues, cuando la variable dependiente representa el intervalo de tiempo entre un “evento inicial” y un “evento terminal” (el cual sólo puede ocurrir una vez). El análisis de supervivencia se asocia con las tablas de mortalidad o tablas de vida para las que el evento final o terminal es el fallecimiento. Es, pues, en el campo de la medicina en donde originariamente se desarrolla esta técnica.

Una de los inconvenientes a los que debe hacer frente el análisis de supervivencia reside en el hecho de que el origen en el tiempo no sea el mismo para todos los individuos. Esta circunstancia provoca que para determinados individuos (observaciones) la información disponible no esté completa.

Pese a estas dificultades, el análisis de supervivencia tal y como se expone en este capítulo, permite la inclusión, y posterior análisis, de datos incompletos siendo imprescindible diferenciar “tipos” de observaciones.

- En primer lugar, hay que distinguir entre datos *censados* (no se ha producido el evento terminal) y *fallecimientos* (sí se ha producido el evento final). Mientras sí se conoce exactamente el tiempo de supervivencia para los fallecimientos se desconoce el de los datos censados.
- A su vez, dentro de las observaciones censadas hay que distinguir entre: *supervivencia* (el evento terminal no se ha dado mientras ha durado la observación); y *abandono* (*withdrawn*) o *mortalidad experimental* (producida por causas ajenas a la investigación). Para los supervivientes su tiempo de supervivencia o de participación equivale a

la duración entre el evento inicial (primera observación) y el período final de observación. Su tiempo de supervivencia nunca se conocerá de forma exacta, sólo podremos afirmar que han tenido, como mínimo, una cierta duración. En los abandonos el tiempo de participación comprende desde el origen hasta la última observación.

Dado que la finalidad de esta técnica es la estimación de la probabilidad de supervivencia en función de la variable tiempo, los momentos de observación más importantes en el análisis de supervivencia son: (1) el origen (momento en que un individuo es incluido en el análisis); (2) última observación (momento en que un individuo abandona o se le pierde la pista por causas ajenas a la investigación); (3) evento terminal o fallecimiento; y (4) cierre del estudio (momento en que finaliza el período de observación para proceder al análisis de datos).

La **función de supervivencia** puede ser calculada siguiendo distintos procedimientos. La elección de uno u otro dependerá de los objetivos e intereses del investigador pues cada uno de ellos contempla particularidades propias y, en última instancia, de las características de los datos disponibles. Así, pues, es posible diferenciar entre:

- Los **métodos Actuarial** y de **Kaplan-Meier** que persiguen obtener mediante probabilidades condicionadas una descripción más precisa de la evolución de la supervivencia a lo largo del tiempo. El método actuarial es más indicado cuando el número de observaciones es elevado. Por su parte, elegiremos el método de Kaplan-Meier cuando al tener un número reducido de observaciones corremos el riesgo de que alguno de los intervalos en los que se dividió el rango de tiempo queden sin ninguna observación o vacíos.
- La **regresión de Cox**, a diferencia de los dos métodos anteriores, parte del conocimiento de que existe un conjunto de variables independientes (cuantitativas o cualitativas) cuyos valores influyen en el tiempo que transcurre hasta que ocurre el suceso final. Su propósito no es otro que el de obtener una función lineal de las variables independientes a partir de la cual estimar, en función del tiempo, la probabilidad de que ocurra el suceso fatal.

Nosotros vamos a aplicar aquí el método actuarial y a él haremos referencia en la exposición del ejemplo que incluimos. Sin embargo, debemos recordar, que la aplicación del método de Kaplan-Meier guarda cierta similitud y que el cálculo de la regresión de Cox lo reservaremos en aquellas circunstancias en las que tengamos constancia de una relación de dependencia. También queremos insistir en que los conceptos e ideas que aquí se presentan ya deben ser familiares para aquellos lectores con conocimientos de análisis demográfico. La tabla de vida o de mortalidad que recogemos en este capítulo no difiere de la expuesta en análisis demográfico con la salvedad de que aquí se presenta el proceso a seguir cuando en su resolución se aplica el paquete estadístico SPSS.

En concreto, los aspectos que vamos a ver en este capítulo son:

1. Presentación de la **tabla de vida o tabla de mortalidad**. En ella podemos diferenciar, perfectamente, dos bloques de información. Las primeras columnas de la misma recogen los principales conceptos relativos a las observaciones y que a modo de presentación hemos relacionado en la introducción del capítulo. A partir de esta información se estiman propiamente las funciones de supervivencia. Les acompañan sus correspondientes errores estándar de los valores estimados para cada una de las tres funciones. Los errores serán interpretados como el error cometido en la estimación y, en consecuencia, serán indicadores de la precisión de las estimaciones. La tabla concluye ofreciendo el índice de supervivencia o tiempo que se espera transcurra hasta que se produzca el fenómeno final.
2. Si bien en sí misma la concreción de una tabla de vida ya ofrece mucha información, ésta se amplía y completa cuando realizamos comparaciones entre grupos a través de sus correspondientes índices de supervivencia. Por ello, nos ha parecido oportuno incluir en segundo lugar un estudio comparativo entre grupos con la tabla de supervivencia.
3. En último lugar, se relacionan los pasos y secuencias a seguir cuando es el paquete estadístico SPSS el aplicado en el análisis de tablas de vida con el método actuarial.

## 2. Elaboración de la Tabla de Vida: funciones de supervivencia

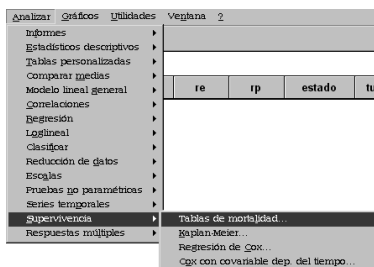


Figura 1

Una vez que hemos decidido que es el método actuarial el que se va a aplicar para el desarrollo del análisis de supervivencia lo primero que se realiza es la tabla de mortalidad o tabla de vida. Esta tabla contiene distintas informaciones descriptivas sobre la evolución de las observaciones.

**1º paso:** Para acceder al cuadro de diálogo que nos permita elaborar estas tablas se sigue la secuencia **Analizar: Supervivencia: Tablas de mortalidad** (figura 1).

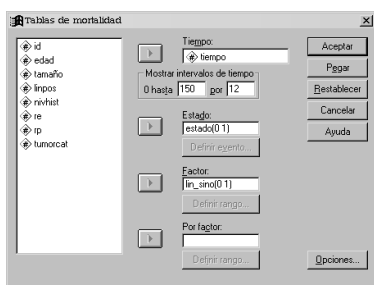


Figura 2

**2º paso:** Una vez en el cuadro de diálogo principal de tablas de mortalidad debemos especificar: la variable **Tiempo** estableciendo los **Intervalos**; la variable **Estado**, que determina si la persona que estamos analizando sigue viva (0) o ha fallecido (1); y por último, introducimos la variable **Factor** que nos permitirá hacer comparaciones de grupos, estableciendo el rango de esa comparación que en nuestro ejemplo es (0=no, 1=si) (figura 2, 2a y 2b).

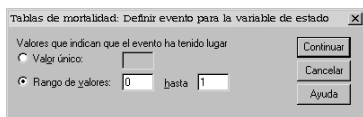


Figura 2a

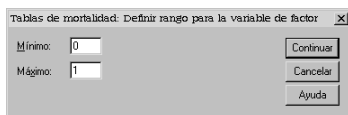


Figura 2b

**3º paso:** Una vez seleccionadas las variables que van a participar en el análisis así como indicado el rango de los intervalos temporales debemos solicitar la realización de la tabla de mortalidad. Para ello cliquearemos sobre el botón de comando **Opciones** que aparece en la parte inferior del cuadro de diálogo principal de **Tablas de mortalidad** seleccionando dicha opción (figura 3). Dentro de este subcuadro y con el objeto de obtener una representación gráfica de la función de supervivencia (una de las series de la tabla de mortalidad) seleccionaremos **Gráfico de Supervivencia**. Por último, y también dentro de este subcuadro, también seleccionaremos la opción **Global** en **Comparar los niveles del primer factor** (los resultados de esta selección nos permitirán establecer comparaciones entre grupos. Ver sección en este capítulo)

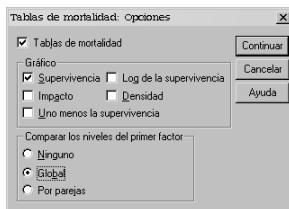


Figura 3

A partir de estas especificaciones obtenemos dos tablas de vida. En la primera nos encontramos con series de información general y que siguen la siguiente estructura:

**1. Tiempo inicial del intervalo (INTVL START TIME).** En esta columna aparece el límite inferior de cada intervalo de tiempo. Por ejemplo, el primer intervalo del caso que aparece en la

sección de resultados (0,0) va de 0 a 12 meses, el segundo (12,0) de 12 a 24, etc. La serie, en su conjunto, nos indica el rango de tiempo en el cual se ha producido el estudio de las observaciones desde un evento inicial hasta un evento final.

**2. Número de entradas en este intervalo** (NUMBER ENTRNG THIS INTVL). Esta columna indica para cada intervalo de tiempo el número total de individuos observados o supervivientes de la serie. En el ejemplo iniciaron la experiencia 929 sujetos de los cuales, en el segundo intervalo, sólo nos quedan 831. Es decir, 831 individuos iniciaron el segundo intervalo de observación. En el último intervalo sólo quedan 3 individuos/observaciones que no han fallecido. Estos son los auténticos supervivientes del fenómeno analizado.

**3. Abandonos** (NUMBER WDRAWN DURING INTVL). En esta serie se recogen el número de individuos que, por causas distintas al fallecimiento, no han llegado al final de período; esto es han abandonado el estudio. En nuestro ejemplo puede apreciarse como no se produce ningún abandono (en todos los intervalos aparece un 0): toda la muestra que se somete a estudio acaba el período de observación.

**4. Número de observaciones expuestas a riesgo** (NUMBER EXPOSED TO RISK). Esta columna se obtiene a partir de una estimación: número de entradas menos la mitad de los abandonos. Dado que en nuestro caso no ha habido abandonos esta columna coincide con la segunda o número de entradas en este intervalo.

**5. Número de eventos terminales** (NUMBER OF TERMNL EVENTS). Esta columna recoge lo que en análisis demográfico se denomina *suceso*. Esta serie nos indica, para cada intervalo de tiempo, el número total fallecimientos o de observaciones que han sido alcanzadas por el suceso fatal.

A partir de estas cinco columnas en las que se recogen en valores absolutos las cifras relacionadas con los distintos conceptos de la tabla de supervivencia y como previo paso al cálculo de las funciones de supervivencia, los datos se relativizan a partir del cálculo de dos series más. Sólo así, recordemos, se pueden efectuar comparaciones.

6. **Proporción de eventos terminales** (PROPN TERMINATING). Esta columna nos indica, para cada intervalo la proporción, de individuos que fallecen.

7. **Proporción de supervivencia** (PROPN SURVIVING). Esta columna complementa a la anterior pues nos dice la proporción de sucesos supervivientes en cada intervalo. La suma de ambas columnas da 1.

La segunda tabla de vida que se desprende una vez aplicado este análisis recoge específicamente las tres funciones de supervivencia (survival functions). La finalidad de estas tres funciones no es otra que la de facilitar la interpretación de los resultados y su cálculo solo es posible una vez halladas las columnas citadas. Estas tres funciones, además, son matemáticamente equivalentes: conocido el valor de una de ellas es posible determinar el de las restantes funciones

8. **Proporción de supervivencia acumulada al final del intervalo** (CUMUL PROPN SURV AT END). Esta columna nos indica la probabilidad de que no sean alcanzados los individuos que pertenecen a un intervalo por el suceso fatal. En nuestro caso la probabilidad sufre un descenso lento y gradual pero ya en el intervalo 36 la probabilidad de que suceda se ha reducido prácticamente a la mitad. Para su cálculo se multiplica la proporción de supervivencia del intervalo anterior por la proporción terminal del intervalo en cuestión. En el primer intervalo se considera que la proporción de supervivencia acumulada del intervalo anterior es igual a uno. Por eso tenemos:  $1 * 0,8945 = 0,8945$ ; para el segundo intervalo  $0,8945 * 0,8315 = 0,7438$ ; para el tercero:  $0,7438 * 0,8191 = 0,6093$ ; y así, sucesivamente.

9. **Probabilidad de densidad** (PROBABILITY DENSITY). Por su parte, esta columna recoge la probabilidad estimada, por unidad de tiempo, del evento terminal que ocurre en el intervalo en cuestión.

10. **La proporción de azar** (HAZARD RATE). Esta última serie proporciona una estimación de la probabilidad, por unidad de tiempo, de que los casos que entran en el intervalo experimentarán el evento terminal en este intervalo.

11. En la parte inferior de esta primera tabla aparece la **mediana del tiempo de supervivencia** (median survival time) y



que en nuestro ejemplo es igual a 43,96 meses. Nótese que la mediana se sitúa en el intervalo para el que la probabilidad de supervivencia se había reducido a la mitad.

Las últimas tres columnas de la tabla de vida corresponden a: (12) los **errores estándar de supervivencia acumulada** (SE OF CUMUL SURVIVING); (13) los **errores de la probabilidad de densidad** (SE OF PROBABILITY DENS); y (14), los **errores de la proporción de azar** (SE OF HAZARD RATE). Los errores estándar de los valores estimados para cada una de las tres funciones estimadas son considerados como una medida de error cometida en la estimación respectivamente. En nuestro caso concreto y considerando que los errores estándar cometidos en la estimación de la función de supervivencia (SE OF CUMUL SURVIVING) son pequeños, podemos concluir en que las estimaciones son bastantes precisas.

### 3. Comparación de la supervivencia en distintos grupos: las puntuaciones de supervivencia

Ya hemos comentado en la introducción del capítulo como además de poder realizar la tabla de vida podemos someter a comparación, bajo los términos de supervivencia, distintos grupos. Para ello es imprescindible disponer de las puntuaciones de supervivencia (*survival score*); éstas se calculan para cada individuo comparando su tiempo de supervivencia con todas las demás observaciones.

En este sentido, las puntuaciones de supervivencia se utilizan en el cálculo del estadístico que, asociado a la correspondiente prueba, nos va a permitir contrastar la hipótesis nula de que la distribución de la supervivencia es la misma en los grupos que deseamos comparar. Si el *p-valor* asociado al estadístico es menor que el nivel de significación (normalmente, 0.05) rechazaremos que la supervivencia es la misma entre los grupos.

Nuestro análisis concluye con la comparación de los grupos, para establecer en nuestro ejemplo, si en función de la información de una variable independiente los datos que se obtienen para un grupo difieren de los obtenidos en el otro grupo. En nuestro caso, queremos saber si la supervivencia entre las perso-

nas a las que se les observa Nodos Linfáticos es igual que a las personas a los que no se les ha observado esta patología.

En las tablas de mortalidad se aparecen para los dos grupos en el anexo de resultados de este capítulo puede apreciarse como la mediana para el primer grupo (sin nodos linfáticos) es mayor que en el grupo con nodos linfáticos. Sin embargo, y considerando la información que en términos comparativos ofrece este indicador, podemos plantearnos si la diferencia que se aprecia es estadísticamente significativa o si de lo contrario responde al azar. Para verificar el supuesto de igualdad planteamos la hipótesis nula de que los dos grupos analizados no presenten diferencias respecto al tiempo que tardan en llegar al suceso final. Para ello, recordemos que en el **3º paso (figura 3)** ya habíamos seleccionado las opciones necesarias para poder llevar a cabo las comparaciones entre grupos.

Para comparar grupos en el análisis de supervivencia se utiliza el **estadístico D** que se calcula a partir de las puntuaciones de supervivencia utilizando el algoritmo de Lee y Desu. La D se distribuye asintóticamente como una ji-cuadrado con  $k - 1$  grados de libertad, siendo  $k$  el número de grupos. La hipótesis nula sostiene que los grupos proceden de la misma población. Si la ji-cuadrado, derivada de la D, supera el intervalo de confianza se rechaza la hipótesis nula con un grado de significación  $p$  (en la parte inferior de la tabla está la prueba de significación).

En el ejemplo tenemos un estadístico D igual a 1,475, con una  $p=0'2246$ . Por lo tanto, podemos aceptar la hipótesis nula. Es decir, a las personas a los que se les observó nodos linfáticos el fallecimiento le sobreviene igual que a los que no se les observó este tipo de patología. El grado de significación ( $p$ ) establece la probabilidad de error al rechazar la hipótesis nula ( $H_0$ : que las dos muestras provengan de la misma población, o sea, que exista diferencias). Por tanto, para poder rechazar la hipótesis nula, tendríamos que obtener un nivel de significación ( $p$ ) por debajo de 0,05, que no es nuestro caso.

Por último, conviene advertir que en el supuesto caso de que hubiéramos rechazado la igualdad en una comparación en la que aparecían implicados más de dos grupos (en nuestro ejemplo es el grupo Si o el grupo No), deberíamos haber comprobado si para cada par de grupos (PAIRWISE COMPARISON) las diferencias también son significativas; esto es, si también podíamos

rechazar la  $H_0$  de igualdad entre pares de grupos. Para cada par de grupos el p-valor debe ser inferior a 0.05. Este análisis se solicita, también, en el subcuadro de diálogo de Opciones de la Tabla de mortalidad. En este caso habría que añadir, junto a la opción de Global, la de **Por Parejas** en la ventana de Comparar los niveles del primer factor.

#### 4. Resultados

---

Los resultados que obtenemos una vez ejecutado un análisis de mortalidad siguiendo el método actuarial con el paquete estadístico SPSS se estructura como sigue:

- En primer lugar, aparecen tantas **tablas de mortalidad** como grupos incluidos en el análisis. En ella cobran especial protagonismo: las series de funciones de supervivencia; la edad mediana en la que acontece el suceso fatal (puede servirnos como primer indicador comparativo); y los errores estándar de los valores estimados para cada una de las funciones de supervivencia (indicadores de la precisión de las estimaciones).
- En segundo lugar, y siempre que estemos comparando dos o más grupos, aparecerá una tabla con la **prueba de significación estadística**.
- Por último, el **gráfico de supervivencia** nos sirve para interpretar y comparar las funciones de supervivencia obtenidas.

#### 4.1. Tablas de Vida: Grupo sin nodos linfáticos

Life Table

Survival Variable TIEMPO Tiempo (meses)  
for LIN\_SINO ¿Nodos Linfáticos? = 0 No

Intrvl Start Time	Number Entng this Intrvl	Number Withdrawn During Intrvl	Number Exposed to Risk	Number of Terml Events	Progn Terml- nating	Progn Sur- viving	Cumul Surv at End	Proba- bilty Densty	Hazard Rate
,0	929,0	,0	929,0	98,0	,1055	,8945	,8945	,0088	,0093
12,0	831,0	,0	831,0	140,0	,1685	,8315	,7438	,0126	,0153
24,0	691,0	,0	691,0	125,0	,1809	,8191	,6093	,0112	,0166
36,0	566,0	,0	566,0	153,0	,2703	,7297	,4446	,0137	,0260
48,0	413,0	,0	413,0	129,0	,3123	,6877	,3057	,0116	,0308
60,0	284,0	,0	284,0	94,0	,3310	,6690	,2045	,0084	,0331
72,0	190,0	,0	190,0	75,0	,3947	,6053	,1238	,0067	,0410
84,0	115,0	,0	115,0	50,0	,4348	,5652	,0700	,0045	,0463
96,0	65,0	,0	65,0	31,0	,4769	,5231	,0366	,0028	,0522
108,0	34,0	,0	34,0	18,0	,5294	,4706	,0172	,0016	,0600
120,0	16,0	,0	16,0	13,0	,8125	,1875	,0032	,0012	,1140
132,0	3,0	,0	3,0	3,0	1,0000	,0000	,0000	,0003	,1667

The median survival time for these data is **43,96 meses**

#### 4.2. Funciones de Supervivencia grupo sin nodos

Intrvl Start Time	SE of Cumul Sur- viving	SE of Proba- bilty Densty	SE of Hazard Rate
,0	,0101	,0008	,0009
12,0	,0143	,0010	,0013
24,0	,0160	,0009	,0015
36,0	,0163	,0010	,0021
48,0	,0151	,0009	,0027
60,0	,0132	,0008	,0033
72,0	,0108	,0007	,0046
84,0	,0084	,0006	,0063
96,0	,0062	,0005	,0089
108,0	,0043	,0004	,0132
120,0	,0019	,0003	,0231
132,0	,0000	,0002	,0000

### 4.3. Tablas de Vida: Grupo con nodos linfáticos

Life Table

Survival Variable		TIEMPO		Tiempo (meses)							
		for		LIN_SINO		¿Nodos Linfáticos?		=		1 Si	
Intrvl Start Time	Number Entng this Intrvl	Number Withdrawn During Intrvl	Number Exposed to Risk	Number of Termpl Events	Progn Termi- nating	Progn Surv- viving	Cumul Progn Surv at End	Proba- bility Densty	Hazard Rate		
	,0	278,0	,0	278,0	33,0	,1187	,8813	,8813	,0099	,0105	
12,0	245,0	,0	245,0	58,0	,2367	,7633	,6727	,0174	,0224		
24,0	187,0	,0	187,0	36,0	,1925	,8075	,5432	,0108	,0178		
36,0	151,0	,0	151,0	33,0	,2185	,7815	,4245	,0099	,0204		
48,0	118,0	,0	118,0	32,0	,2712	,7288	,3094	,0096	,0261		
60,0	86,0	,0	86,0	32,0	,3721	,6279	,1942	,0096	,0381		
72,0	54,0	,0	54,0	23,0	,4259	,5741	,1115	,0069	,0451		
84,0	31,0	,0	31,0	9,0	,2903	,7097	,0791	,0027	,0283		
96,0	22,0	,0	22,0	9,0	,4091	,5909	,0468	,0027	,0429		
108,0	13,0	,0	13,0	7,0	,5385	,4615	,0216	,0021	,0614		
120,0	6,0	,0	6,0	6,0	1,0000	,0000	,0000	,0018	,1667		

#### 4.5. Significación del Análisis

Comparison of survival experience using the Wilcoxon (Gehan) statistic

Survival Variable	TIEMPO	Tiempo (meses)					
grouped by	LIN_SINO	¿Nodos Linfáticos?					

Overall comparison	<b>statistic</b>	<b>1,475</b>	D.F.	1	<b>Prob.</b>	<b>,2246</b>
--------------------	------------------	--------------	------	---	--------------	--------------

Group	label	Total N	Uncen	Cen	Pct Cen	Mean Score
0	No	929	929	0	,00	13,3305
1	Sí	278	278	0	,00	-44,5468

#### 4.6. Representación Gráfica de los dos grupos

